AD-A189 209 THE EFFECT OF SOFTWARE SIZE UNCERTAINTY ON EFFORT 1/1
ESTIMATES GENERALIZED B.. (U) MITRE CORP BEDFORD MA
P R GARVEY NOV 87 MTR-10212 ESD-TR-87-169
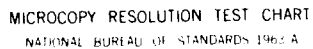
UNCLASSIFIED F19628-86-C-0001 F/G 12/5 NL

END
DATE
FILMED
5 88

MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963 A

ESD-TR-87-169

MTR-10212

THE EFFECT OF SOFTWARE SIZE UNCERTAINTY ON EFFORT ESTIMATES GENERATED BY αI^3 SOFTWARE RESOURCE MODELS

BY

PAUL R. GARVEY

NOVEMBER 1987

AD-A189 209

Prepared for

DEPUTY COMMANDER FOR STRATEGIC SYSTEMS
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
Hanscom Air Force Base, Massachusetts

DTIC
ELECTE
FEB 0 1 1988
H

88 1 27 168

REVIEW AND APPROVAL

This technical report has been reviewed and is approved for publication.

RICHARD C. PASCHALL, JR., COLONEL, USAF
System Program Director
North Atlantic Defense Systems Directorate

FOR THE COMMANDER

LARRY W. CURL, COLONEL, USAF
Director, Management Services
Deputy Commander for Strategic Systems

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release, distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	

4. PERFORMING ORGANIZATION REPORT NUMBER(S) MTR-10212 ESD-TR-87-169	5. MONITORING ORGANIZATION REPORT NUMBER(S)

6a. NAME OF PERFORMING ORGANIZATION The MITRE Corporation	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION

6c. ADDRESS (City, State, and ZIP Code) Burlington Road Bedford, Massachusetts 01730	7b. ADDRESS (City, State, and ZIP Code)

8a. NAME OF FUNDING / SPONSORING ORGANIZATION Deputy Commander for Strategic Systems	8b. OFFICE SYMBOL (If applicable) XYBX	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F19628-86-C-0001

8c. ADDRESS (City, State, and ZIP Code) Electronic Systems Division, AFSC Hanscom AFB, Massachusetts 01731-5000	10. SOURCE OF FUNDING NUMBERS			
	PROGRAM ELEMENT NO.	PROJECT NO. 443C	TASK NO	WORK UNIT ACCESSION NO

11. TITLE (Include Security Classification)

The Effect of Software Size Uncertainty on Effort Estimates Generated by αI^β Software Resource Models

12. PERSONAL AUTHOR(S)

Garvey, Paul R.

13a. TYPE OF REPORT Final	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Year, Month, Day) 1987 November	15. PAGE COUNT 41

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	CSCI Size Measuring Effort Uncertainty Software Resource Models

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

In support of the Cost Technical Center's estimate of the North Atlantic Defense System (NADS) software development effort, a new probability model, the Analytic Software Effort Probability (ASEP) model, has been developed that quantifies the impact of software size uncertainty on effort estimates generated by αI^β software resource models. This paper provides the mathematical foundation of the ASEP model.

Analytic equations that yield the exact density, cumulative distribution, mean, and variance of the development effort, given that the size of a single Computer Software Configuration Item (CSCI) is a uniform or triangularly distributed random variable, are described. These equations are extended to the case of approximating the overall development effort mean, variance, and distribution for a system consisting of multiple CSCIs.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Diana F. Arimento	22b. TELEPHONE (Include Area Code) (617)271-7454	22c. OFFICE SYMBOL Mail Stop D230

DD Form 1473, JUN 86   Previous editions are obsolete.   SECURITY CLASSIFICATION OF THIS PAGE

ACKNOWLEDGMENTS

QUALITY INSPECTED 2

Accession For	
NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

iii

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

SECTION 1

INTRODUCTION

1.1 BACKGROUND AND PURPOSE

The general mathematical form of many software resource models
is reducible to the expression

$$DE = aI^{\beta} \Pi f_j \qquad (j = 1, \ldots, n) \text{ and } a, \beta > 0 \qquad (1)$$

where DE is the development effort measured in staff months, a and β
are constants specific to the model type, I denotes the size of the
Computer Software Configuration Item (CSCI) measured in thousands of
delivered source instructions (KDSI), and Πf_j is the product of n
development effort multipliers which depict unique attributes of the
software product, host computer hardware, programmer and analyst
experience, and product schedule. If the size I of a CSCI is a
random variable with a given probability distribution, then DE is a
random variable; and closed form expressions for the probability
distribution of DE and the mean and variance of DE may be derived
for selected probability distributions.

The purpose of this paper is to provide the closed form
equations for the density and distribution of DE and the mean and
variance of DE for a set of prescribed size probability distribu-
tions. These equations are needed for determining analytically the
effort uncertainty due to the size uncertainty associated with the
development of the Computer Software Configuration Item.

This analysis does not preclude the fact that there exist other
technical and programmatic uncertainties which impact the develop-
ment of software. Multiple site development, subcontracting,
personnel experience, concurrent hardware development, and the
availability of a sound programming support environment are

1

additional sources of uncertainty which should be quantified when a formal software risk analysis is performed.

1.2 SCOPE OF THE PAPER

Two classes of size probability distribution functions will be defined for this analysis. They are the:

- o Uniform Distribution
- o Triangular Distribution

Within their permissable intervals of definition, the densities of these size probability distribution functions are characterized by the common feature of unimodality and finite range. In general, these densities are typical of the shapes encountered when size assessments for a CSCI are made.

For each size probability distribution function described above, Section 2 provides equations which define the exact development effort probability distribution and the exact effort mean and variance specific to that CSCI. Section 3 presents a graphical representation of the effort probability distributions which arise when the size of a CSCI is transformed to the unit interval. This is useful for determining the effort associated with a specified level of risk. Section 4 introduces the Analytic Software Effort Probability (ASEP) model. The ASEP model is an extension of the equations developed in Section 2 for the purpose of approximating the development effort mean, variance, and distribution for a system consisting of multiple CSCIs. Appendix A provides case studies which describe how the ASEP model is applied to single and multiple CSCI systems.

2

SECTION 2

MEASURING EFFORT UNCERTAINTY IN αI^{β} SOFTWARE
RESOURCE MODELS – AN ANALYTIC APPROACH

2.1 INTRODUCTION

There are a number of software resource models which are reducible to the form shown in equation 1. Examples of such models are the Constructive Cost Model (COCOMO) [1], the Jensen Model [2], the Doty Model [5], and the Walston-Felix Model [5]. It is not within the scope of this paper to discuss the details of these software resource models. The reader is directed to the references if such information is desired. The mathematical analysis presented in this paper is focused on determining analytical expressions for the probability distribution of DE and the mean and variance of DE, given that the size I of a CSCI is a random variable from one of the classes of size probability distributions described in Section 2.2. The analysis methodology is general to any software resource model of the form

$$DE = \alpha I^{\beta} \qquad \alpha, \beta > 0 \qquad (1)$$
$$\alpha = a \, \Pi f_{j} \qquad (j = 1, \ldots, n)$$

where DE is the development effort measured in staff months, I denotes the size of the CSCI measured in thousands of delivered source instructions, a and β are constants specific to the model type, and Πf_{j} is the product of n development effort multipliers which describe unique characteristics of the software product, host computer hardware, programmer and analyst experience, and product schedule.

3

2.2 SELECTED CLASSES OF SIZE PROBABILITY DISTRIBUTIONS

Given that the size I of a CSCI is a random variable, two classes of size probability distribution functions are discussed in this section. They are the:

- o Uniform Distribution
- o Triangular Distribution

Other probability distribution functions such as the beta distribution may be used to define the size uncertainty of a CSCI, however, the mathematical analysis in this paper quickly becomes unwieldy if size is beta distributed. The uniform and triangular distributions may be considered robust in the sense that they describe reasonably well a wide variety of size distributions, even though they may not always be the most precise representation of the size uncertainty for a particular CSCI. Furthermore, given the degree of uncertainty often attributed to size estimates, and the subjectivity by which a size distribution is ascribed to that CSCI, the uniform and triangular distribution functions can be adequately defined in order to bound this uncertainty. One approach to bounding this uncertainty is to have the size estimator determine, not an absolute size range, but an 80th, or 90th percentile confidence interval from which the analyst may then compute the absolute extremes of the size interval.

The distribution function for I, denoted by $F_I(t)$, is a function of a real variable t such that

$$F_I(t) = \text{Prob}(I \leq t) = \int_{-\infty}^{t} f_I(u)du$$

where f_I is the probability density function for size I.

Assuming that the software development effort DE is given by

$$DE = \alpha I^{\beta} \triangleq g(I) \tag{2}$$

where $I \geq 0$ and $\alpha, \beta > 0$ then $g(I)$ is a strictly monotonic function, and $g^{-1}(I)$ exists uniquely. Therefore, the probability distribution function of DE, denoted by F_{DE}, is given in equation 3.

$$F_{DE}(x) = F_I(g^{-1}(x)) \tag{3}$$

This is true since g is a strictly monotonic increasing function, thus

$$F_{DE}(x) = Prob(DE \leq x) \quad \text{by definition}$$

$$= Prob(g(I) \leq x)$$

$$= Prob(I \leq g^{-1}(x))$$

$$= F_I(g^{-1}(x))$$

where $g(I) \leq x$ if and only if $I \leq g^{-1}(x)$. Since g is a strictly monotonic differentiable function, then the density of DE, denoted by f_{DE}, is given below in equation 4.

$$f_{DE}(x) = f_I(g^{-1}(x)) \, d(g^{-1}(x))/dx \tag{4}$$

Equation 5 defines the expected value of the development effort, and is denoted by $E(DE)$.

$$E(DE) = E(g(I)) = \int_{-\infty}^{\infty} g(u) f_I(u) du \tag{5}$$

5

The effort variance σ^2_{DE} is defined by

$$\sigma^2_{DE} = E((g(I) - \mu_{g(I)})^2) = \alpha^2 E(I^{2\beta}) - (E(DE))^2 \qquad (6)$$

where

$$E(I^{2\beta}) = \int_{-\infty}^{\infty} u^{2\beta} f_I(u) du \qquad (7)$$

The following subsections describe the size probability density $f_I(t)$ and distribution functions $F_I(t)$ for each of the two classes considered in this paper. Equations for the effort probability distribution, the effort density, the effort mean and variance for each of the size probability distributions will be provided in Section 2.3.

2.2.1 The Uniform Distribution

The simplest possible continuous random variable is the uniform random variable. If size I is uniformly distributed on the interval $t_a < t < t_b$, that is

$$I \sim \text{Unif}(t_a, t_b)$$

then the probability density function of I, denoted by f_I, is

$$f_I(t) = \begin{cases} 1/(t_b - t_a) & \text{if } t_a < t < t_b \\ 0 \text{ otherwise} \end{cases} \qquad (8)$$

The graph of f_I is shown in figure 1.

6
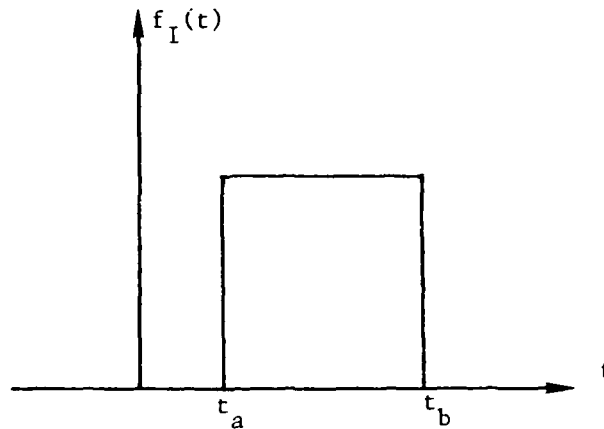
Figure 1

The Uniform Density Function

The probability distribution function of I, assuming that I is uniformly distributed on the interval $t_a \leq t \leq t_b$, is given below in equation 9.

$$F_I(t) = \begin{cases} 0 \text{ if } t < t_a & (9) \\ (t-t_a)/(t_b-t_a) & \text{if } t_a \leq t < t_b \\ 1 \text{ if } t \geq t_b \end{cases}$$

2.2.2 The Triangular Distribution

If the size range of a CSCI can be expressed by three points t_a, t_m, t_b where t_a is the lowest possible size estimate, t_m is the modal size estimate (informally referred to as the most likely), and t_b is the highest possible size estimate, then a triangular density function may be suitable for describing the size distribution of the Computer Software Configuration Item (CSCI). Therefore, if I is triangularly distributed, that is:

$$I \sim Trng\ (t_a,\ t_m,\ t_b)$$

7

where $t_a \leq t_m \leq t_b$, then the probability density function of I, denoted by f_I, is

$$f_I(t) = \begin{cases} (c/(t_m-t_a))(t-t_a) & \text{if } t_a < t \leq t_m \\[2ex] (c/(t_m-t_b))(t-t_b) & \text{if } t_m < t \leq t_b \end{cases} \qquad (10)$$

where $c = 2/(t_b-t_a)$

The graph of f_I is shown in figure 2.
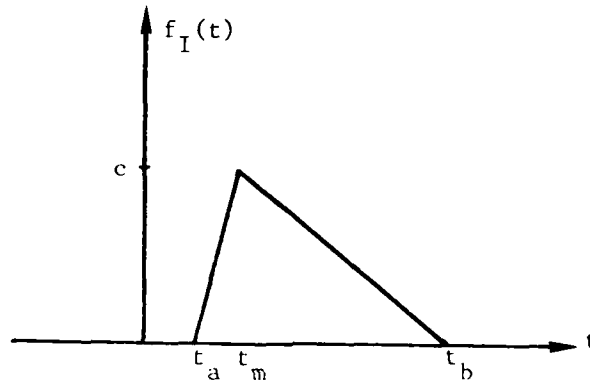
Figure 2

The Triangular Density Function

The probability distribution function of I, assuming that I is triangularly distributed on the interval $t_a < t_m < t_b$ is given below in equation 11.

$$F_I(t) = \begin{cases} 0 & \text{if } t < t_a \\[2ex] (t_b-t_a)^{-1}(t_m-t_a)^{-1}(t-t_a)^2 & \text{if } t_a \leq t < t_m \qquad (11) \\[2ex] 1+(t_b-t_a)^{-1}(t_m-t_b)^{-1}(t-t_b)^2 & \text{if } t_m \leq t < t_b \\[2ex] 1 & \text{if } t \geq t_b \end{cases}$$

8

2.2.3 The Right Triangular Distribution

The Right Triangular Distribution is a special case of the general triangular distribution described in the previous subsection. If the size range of a CSCI can be characterized by the rule that the low size estimate is equal to the modal size estimate (i.e., $t_a = t_m$) and if there exists only upward size uncertainty given by t_b, then it may be said that I has a right triangular distribution, that is:

$$I \sim RtTrng \ (t_a = t_m, \ t_b)$$

The probability density function for I is given by

$$f_I(t) = \begin{cases} (c/(t_b - t_a))(t_b - t) & \text{if } t_a \leq t \leq t_b \\ \\ 0 & \text{otherwise} \end{cases} \qquad (12)$$

The graph of f_I is shown in figure 3. The probability distribution function for I, denoted by $F_I(t)$, is given below in equation 13.

$$F_I(t) = \begin{cases} 0 & \text{if } t < t_a \\ \\ 1 - (t_b - t)^2/(t_b - t_a)^2 & \text{if } t_a \leq t < t_b \\ \\ 1 & \text{if } t \geq t_b \end{cases} \qquad (13)$$

Since the uncertainty surrounding software size estimates almost always exhibits positively skewed distributions, equations which describe negatively skewed right triangular size distributions (i.e, $t_m = t_b$) have not been developed in this paper.

9
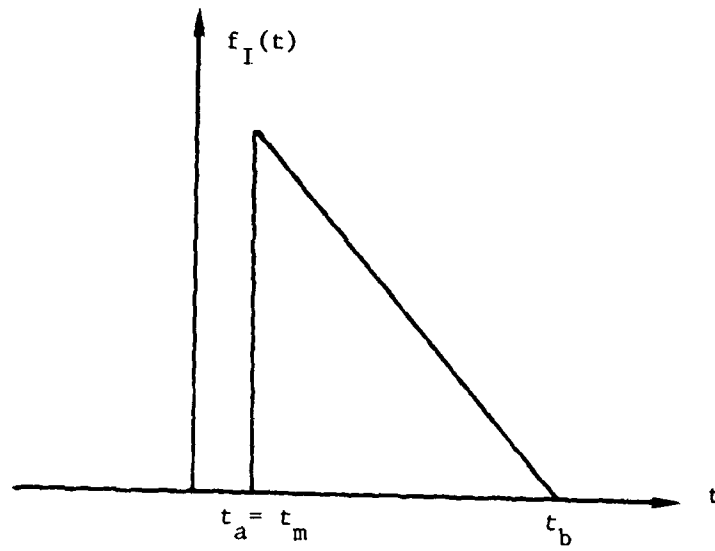
Figure 3

Right Triangular Density Function

10

2.3 COMPUTING THE EFFORT DISTRIBUTION FROM A GIVEN
SIZE PROBABILITY DISTRIBUTION

This section provides the derivation of the mathematical expressions for the density (f_{DE}), the distribution (F_{DE}), the mean ($E(DE)$), and the variance (σ^2_{DE}) of DE given that the distribution of I is triangular. The analysis methodology is the same if I is uniformly or right triangularly distributed. Table 1 provides a general summary of the equations for f_{DE}, F_{DE}, $E(DE)$, and σ^2_{DE} for the cases where I is uniformly, triangularly, or right triangularly distributed.

Given that

$$I \sim \text{Trng} (t_a, t_m, t_b)$$

then from equation 4, the development effort density f_{DE} is

$$f_{DE}(x) = \beta^{-1} x^{-1} (x/\alpha)^{1/\beta} f_I(g^{-1}(x))$$

which may be written as

$$
\begin{cases}
c(t_m - t_a)^{-1} \beta^{-1} x^{-1} (x/\alpha)^{1/\beta} ((x/\alpha)^{1/\beta} - t_a) & \text{if } t_a \leq (x/\alpha)^{1/\beta} < t_m \\
c(t_m - t_b)^{-1} \beta^{-1} x^{-1} (x/\alpha)^{1/\beta} ((x/\alpha)^{1/\beta} - t_b) & \text{if } t_m \leq (x/\alpha)^{1/\beta} < t_b
\end{cases}
\tag{14}
$$

11

From equation 3, the development effort distribution function F_{DE} is

$$\begin{cases} 0 & \text{if } (x/\alpha)^{1/\beta} < t_a \\ (t_b - t_a)^{-1} (t_m - t_a)^{-1} ((x/\alpha)^{1/\beta} - t_a)^2 & \text{if } t_a \leq (x/\alpha)^{1/\beta} < t_m \\ 1 + (t_b - t_a)^{-1} (t_m - t_b)^{-1} ((x/\alpha)^{1/\beta} - t_b)^2 & \text{if } t_m \leq (x/\alpha)^{1/\beta} < t_b \\ 1 & \text{if } (x/\alpha)^{1/\beta} \geq t_b \end{cases} \quad (15)$$

From equation 5, the expected value of the development effort E(DE) is

$$\alpha c (t_m - t_a)^{-1} \left[(t_m^{\beta+2} - t_a^{\beta+2})/(\beta + 2) + (t_a^{\beta+2} - t_a t_m^{\beta+1})/(\beta + 1) \right]$$
$$+ \alpha c (t_m - t_b)^{-1} \left[(t_b^{\beta+2} - t_m^{\beta+2})/(\beta + 2) + (t_b t_m^{\beta+1} - t_b^{\beta+2})/(\beta + 1) \right] \quad (16)$$

and the variance of the development effort σ^2_{DE} based on equation 6 is

$$\sigma^2_{DE} = \alpha^2 E(I^{2\beta}) - (E(DE))^2$$

where $E(I^{2\beta})$ is given by

$$c (t_m - t_a)^{-1} \left[(t_m^{2\beta+2} - t_a^{2\beta+2})/(2\beta + 2) + (t_a^{2\beta+2} - t_a t_m^{2\beta+1})/(2\beta + 1) \right]$$
$$+ c (t_m - t_b)^{-1} \left[(t_b^{2\beta+2} - t_m^{2\beta+2})/(2\beta + 2) + (t_b t_m^{2\beta+1} - t_b^{2\beta+2})/(2\beta + 1) \right] \quad (17)$$

Table 1 summarizes these results for the case where I is triangularly distributed. Furthermore, the equations for f_{DE}, F_{DE}, E(DE), and σ^2_{DE} for the cases where I is uniformly and right triangularly distributed are also provided in table 1.

12

Table 1
A Summary of the Equations
For f_{DE}, F_{DE}, μ_{DE}, and σ^2_{DE}

Distribution of I	Development Effort Density $f_{DE}(x)$	Development Effort Distribution $F_{DE}(x)$
Unif (t_a, t_b)	$\alpha^{-1}\beta^{-1}(x/\alpha)^{(1/\beta)-1}(t_b - t_a)^{-1}$ if $t_a \leq (x/\alpha)^{1/\beta} \leq t_b$	0 if $(x/\alpha)^{1/\beta} < t_a$ $(t_b - t_a)^{-1}((x/\alpha)^{1/\beta} - t_a)$ if $t_a \leq (x/\alpha)^{1/\beta} < t_b$ 1 if $(x/\alpha)^{1/\beta} \geq t_b$
Trng (t_a, t_m, t_b)	$c(t_m - t_a)^{-1}\beta^{-1}x^{-1}(x/\alpha)^{1/\beta}((x/\alpha)^{1/\beta} - t_a)$ if $t_a \leq (x/\alpha)^{1/\beta} < t_m$ $c(t_m - t_b)^{-1}\beta^{-1}x^{-1}(x/\alpha)^{1/\beta}((x/\alpha)^{1/\beta} - t_b)$ if $t_m \leq (x/\alpha)^{1/\beta} < t_b$	0 if $(x/\alpha)^{1/\beta} < t_a$ $(t_b - t_a)^{-1}(t_m - t_a)^{-1}((x/\alpha)^{1/\beta} - t_a)^2$ if $t_a \leq (x/\alpha)^{1/\beta} < t_m$ $1 + (t_b - t_a)^{-1}(t_m - t_b)^{-1}((x/\alpha)^{1/\beta} - t_b)^2$ if $t_m \leq (x/\alpha)^{1/\beta} < t_b$ 1 if $(x/\alpha)^{1/\beta} \geq t_b$
RltTrng $(t_a = t_m, t_b)$	$\beta^{-1}x^{-1}(x/\alpha)^{1/\beta}c(t_b - t_a)^{-1}(t_b - (x/\alpha)^{1/\beta})$ if $t_a \leq (x/\alpha)^{1/\beta} \leq t_b$	0 if $(x/\alpha)^{1/\beta} < t_a$ $1 - (t_b - t_a)^{-1}(t_b - (x/\alpha)^{1/\beta})$ if $t_a \leq (x/\alpha)^{1/\beta} < t_b$ 1 if $(x/\alpha)^{1/\beta} \geq t_b$

Note: x is development effort measured in staff-months.

Distribution of I	Development Effort Mean $E(DE)$	Development Effort Variance σ^2_{DE}
Unif (t_a, t_b)	$E(DE)_{Unif} = \alpha(\beta + 1)^{-1}(t_b - t_a)^{-1}(t_b^{\beta+1} - t_a^{\beta+1})$	$\alpha^2 c(2\beta + 1)^{-1}(t_b - t_a)^{-1}(t_b^{2\beta+1} - t_a^{2\beta+1}) - (E(DE)_{Unif})^2$
Trng (t_a, t_m, t_b)	$E(DE)_{Trng} = \alpha c(t_m - t_a)^{-1}\left[(t_m^{\beta+2} - t_a^{\beta+2})/(\beta + 2) + (t_a^{\beta+2} - t_a t_m^{\beta+1})/(\beta + 1)\right]$ $+ \alpha c(t_m - t_b)^{-1}\left[(t_b^{\beta+2} - t_m^{\beta+2})/(\beta + 2) + (t_b t_m^{\beta+1} - t_b^{\beta+2})/(\beta + 1)\right]$	$\alpha^2 c(t_m - t_a)^{-1}\left[(t_m^{2\beta+2} - t_a^{2\beta+2})/(2\beta + 2) + (t_a^{2\beta+2} - t_a t_m^{2\beta+1})/(2\beta + 1)\right]$ $+ \alpha^2 c(t_m - t_b)^{-1}\left[(t_b^{2\beta+2} - t_m^{2\beta+2})/(2\beta + 2) + (t_b t_m^{2\beta+1} - t_b^{2\beta+2})/(2\beta + 1)\right]$ $- (E(DE)_{Trng})^2$
RltTrng $(t_a = t_m, t_b)$	$E(DE)_{RltTrng} = \alpha c(t_b - t_a)^{-1}\left[(t_a^{\beta+2} - t_b^{\beta+2})/(\beta + 2) + (t_b^{\beta+2} - t_b t_a^{\beta+1})/(\beta + 1)\right]$	$\alpha^2 c(t_b - t_a)^{-1}\left[(t_b^{2\beta+2} - t_b^{2\beta+1})/(2\beta + 1) + (t_a^{2\beta+2} - t_b t_a^{2\beta+1})/(2\beta + 2)\right]$ $- (E(DE)_{RltTrng})^2$

13

SECTION 3

TRANSFORMING THE EFFORT PROBABILITY
DISTRIBUTION TO THE UNIT INTERVAL

This section provides a method whereby the direct computation
of effort at which there is a given level of risk can be made.
Consider the following illustration:

Suppose the development effort DE at the 80th percentile
level of a CSCI whose size is triangularly distributed
with $t_a = 30$, $t_m = 50$, and $t_b = 80$ needs to be computed.
Furthermore, let $\alpha = 2.8$ and $\beta = 1.2$ as used in the COCOMO
embedded mode equation [1]. From equation 1 the computed
development effort DE evaluated at the mode size
(informally referred to as the most likely) $t_m = 50$ is:

$$DE = 2.8 \ (50)^{1.2} = 306 \text{ staff months}$$

Now define Z as

$$0 \leq Z = \text{Prob(DE} \leq 306) = (t_m - t_a)/(t_b - t_a) = 0.4 \leq 1$$

where Z is the probability with which the computed
development effort based on the most likely size will
occur. Notice that Z approaches one as the most likely
size approaches the absolute upperbound size (assumed to
be given by t_b), and Z approaches zero when the most likely
size approaches the absolute lower bound size (assumed to
be given by t_a) The value of Z, therefore, describes how

14

the size distribution is skewed. Figure 4 shows a family
of development effort probability curves which result when
Z varies between zero and one. Since this illustration
yielded Z = 0.4, selecting the curve Z_4 = 0.4 from figure
4, the complete probability distribution of DE is shown.
The development effort at the 80th percentile level along
this curve occurs when n (the X axis in Figure 4) is
approximately 0.6. Equations 20 through 23 (to be derived
in this section) with n = 0.6 yields an 80th percentile
development effort DE of 401 staff months, the desired
result.

The following discussion develops the mathematical basis for
the family of distribution profiles shown in figure 4, and the
equations which determine the development effort DE at any specified
level of risk.

Suppose that

$$I \sim Trng(t_a, t_m, t_b)$$

where t_a, t_m, and t_b define the absolute range of the triangular
probability density function (pdf) shown in figure 2. The size
estimate associated with the mode of this pdf is given by t_m. It is
possible to transform the general triangular pdf shown in figure 2
to the unit size interval through the transformation

$$0 < Z = (t_m - t_a)(t_b - t_a)^{-1} < 1 \qquad (18)$$

where Z is the exact representation of t_m, but expressed as a number
between zero and one. The value of Z reflects the skewness of the

15
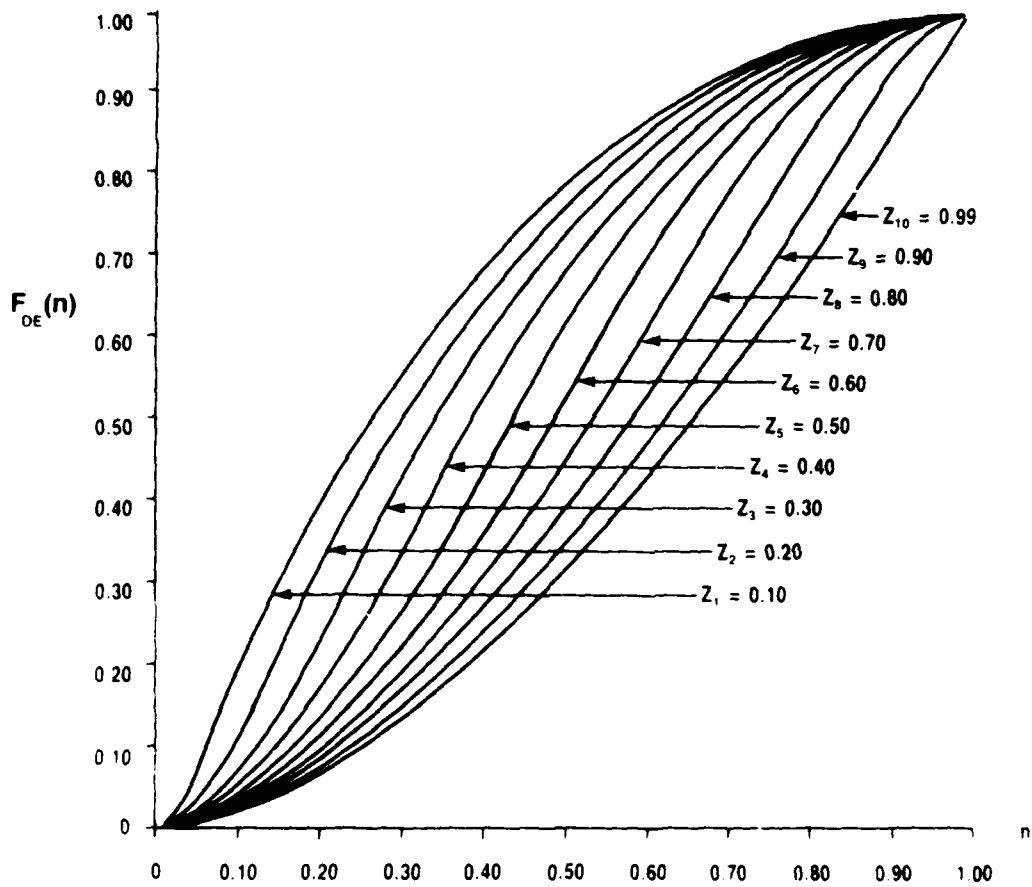
Figure 4
The Distribution Profile of F_{DE}

16

pdf. Furthermore, Z is the probability with which the computed development effort based on the mode size $DE(t_m)$ will occur, that is

$$Prob(DE \leq DE(t_m)) = Z$$

Given that I is mapped into the unit interval via the Z transformation shown in equation 18, define the dimensionless function e by

$$e = \alpha(t - t_a)^\beta (t_b - t_a)^{-\beta} \qquad t_a \leq t \leq t_b \qquad (20)$$

where the development effort DE shown in equation 1 is related to e by

$$DE = \alpha[(e/\alpha)^{1/\beta} (t_b - t_a) + t_a]^\beta \qquad t_a \leq t \leq t_b \qquad (21)$$

Let

$$n = (t - t_a)^\beta (t_b - t_a)^{-\beta} \qquad t_a \leq t \leq t_b \qquad (22)$$

from which

$$e = \alpha n \qquad 0 \leq n \leq 1 \qquad (23)$$

for some n bounded by the unit interval. From table 1 the development effort probability distribution F_{DE} on the unit interval (assuming that size I is triangularly distributed) is

$$F_{DE} = \begin{cases} 0 & \text{if } (n)^{1/\beta} < 0 \\ Z^{-1}(n)^{2/\beta} & \text{if } 0 \leq (n)^{1/\beta} < Z \\ 1+(Z-1)^{-1}[(n)^{1/\beta} -1]^2 & \text{if } Z \leq (n)^{1/\beta} < 1 \\ 1 & \text{if } (n)^{1/\beta} \geq 1 \end{cases} \qquad (24)$$

Equation 24 was obtained from table 1 by setting $t_a = 0$, $t_b = 1$, $t_m = Z$, and $n = x/\alpha$. Table 2 provides the numerical values of F_{DE} for various n.

An exact value of the development effort at a specified level γ (defined by DE_γ) can be obtained by equations 25 and 26. Equations 25 and 26 were obtained by setting F_{DE} equal to γ and solving equation 15 for x.

$$DE_\gamma = \begin{cases} \alpha(t_a + [\gamma(t_b - t_a)(t_m - t_a)]^{1/2})^\beta & \text{if} \quad \gamma < Z & (25) \\ \\ \alpha(t_b - [(1-\gamma)(t_b - t_a)(t_b - t_m)]^{1/2})^\beta & \text{if} \quad \gamma \geq Z & (26) \end{cases}$$

18

Table 2

$F_{DE}(n)$[1] Versus Z Along the Unit Interval

n	z_1 0.10	z_2 0.20	z_3 0.30	z_4 0.40	z_5 0.50	z_6 0.60	z_7 0.70	z_8 0.80	z_9 0.90	z_{10} 0.99
0.01	0.0046	0.0023	0.0015	0.0012	0.0009	0.0008	0.0007	0.0006	0.0005	0.0005
0.02	0.0147	0.0074	0.0049	0.0037	0.0029	0.0025	0.0021	0.0018	0.0016	0.0015
0.05	0.0679	0.0339	0.0226	0.0170	0.0136	0.0113	0.0097	0.0085	0.0075	0.0069
0.07	0.1180	0.0594	0.0396	0.0297	0.0238	0.0198	0.0170	0.0149	0.0132	0.0120
0.09	0.1676	0.0904	0.0602	0.0452	0.0361	0.0301	0.0258	0.0226	0.0201	0.0183
0.10	0.1911	0.1077	0.0718	0.0539	0.0431	0.0359	0.0308	0.0269	0.0239	0.0218
0.20	0.3941	0.3183	0.2280	0.1710	0.1368	0.1140	0.0978	0.0855	0.0760	0.0691
0.30	0.5543	0.4986	0.4270	0.3361	0.2689	0.2241	0.1921	0.1681	0.1494	0.1358
0.40	0.6832	0.6436	0.5926	0.5247	0.4343	0.3620	0.3102	0.2714	0.2413	0.2193
0.50	0.7861	0.7593	0.7250	0.6791	0.6150	0.5250	0.4500	0.3937	0.3500	0.3182
0.60	0.8665	0.8498	0.8283	0.7997	0.7596	0.6995	0.6098	0.5335	0.4743	0.4311
0.70	0.9265	0.9174	0.9056	0.8898	0.8678	0.8347	0.7796	0.6898	0.6132	0.5574
0.80	0.9680	0.9640	0.9589	0.9520	0.9424	0.9280	0.9040	0.8560	0.7660	0.6964
0.90	0.9921	0.9912	0.9899	0.9882	0.9859	0.9823	0.9764	0.9647	0.9293	0.8474
0.99	0.9999	0.9999	0.9999	0.9999	0.9999	0.9998	0.9998	0.9997	0.9993	0.9930

Note 1: $F_{DE}(n)$ is rounded and computed at $\beta = 1.2$.

19

THE ANALYTIC SOFTWARE EFFORT PROBABILITY (ASEP) MODEL

The purpose of this section is to introduce the Analytic Software Effort Probability (ASEP) model. The ASEP model is an extension to the equations developed in Section 2 for the purpose of approximating the development effort mean, variance, and probability distribution for a system consisting of two or more interrelated Computer Software Configuration Items (CSCIs). The Constructive Cost Model (COCOMO) [1] provides the framework around which the Analytic Software Effort Probability (ASEP) model is developed.

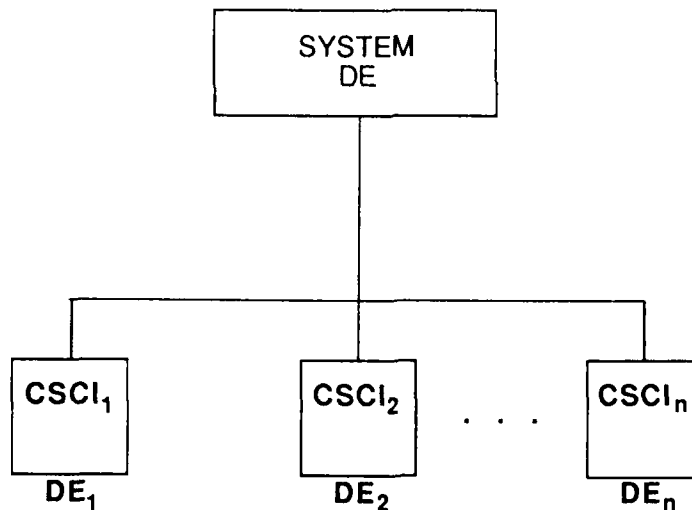Consider the hypothetical n-CSCI system shown in figure 5.

Figure 5

An n-CSCI System

The Central Limit Theorem [3] will be used to contend that the development effort distribution is approximately Gaussian. The Central Limit Theorem states that if X_1, X_2, ..., X_n are independent random variables each with finite mean μ_i and variance σ_i^2, then the distribution of their sum is asymptotically normal, with a mean equal to $\sum \mu_i$ and variance equal to $\sum \sigma_i^2$ for $i = 1$, ..., n. It would be convenient to assert that the total system development effort DE is approximately normal, with mean equal to the sum of the mean efforts for each CSCI, and variance equal to the sum of their variances. However, the Constructive Cost Model applied to a system of interrelated CSCIs exhibits the property that the development efforts DE_1, DE_2, ..., DE_n are not independent random variables. Proposed below is a method of approximating the development effort DE such that DE is expressed as a sum of independent random variables, from which the Central Limit Theorem is then valid.

Suppose that the size uncertainty of each CSCI in the n-CSCI system shown in figure 5 can be characterized as being uniformly or triangularly distributed. The COCOMO nominal development effort relationship is

$$DE = a\left(\sum I_i\right)^\beta \qquad (i = 1, \ldots, n) \qquad (27)$$

The nominal development effort for each CSCI is given by

$$DE_i = (I_i)\left(\sum I_i\right)^{-1} DE$$

$$= (I_i)\left(\sum I_i\right)^{-1} a\left(\sum I_i\right)^\beta$$

$$= a I_i\left(\sum I_i\right)^{\beta-1} \qquad (i = 1, \ldots, n) \qquad (28)$$

21

Clearly, the development efforts DE_i are not independent since a change in the size of one CSCI affects the estimated development efforts of all the other CSCIs in the system. It is possible to find a constant φ which provides a way to express DE as a sum of independent random variables by solving the expression

$$a\left(\sum \bar{I}_i\right)^\beta = \varphi \cdot a \sum \bar{I}_i^\beta \quad (i = 1, \ldots, n) \tag{29}$$

for φ, where $\bar{I}_i$ is the mean size of $CSCI_i$.

Let $\hat{DE}$ represent an approximation to the overall development effort DE given in equation 27. Thus

$$DE \cong \hat{DE} = \varphi \sum \hat{DE}_i = \varphi \sum a I_i^\beta \quad (i = 1, \ldots, n) \tag{30}$$

Thus, DE is now the sum of independent random variables. Therefore, the mean of DE is given by

$$E(DE) \cong \varphi \sum E(\hat{DE}_i) \quad (i = 1, \ldots, n) \tag{31}$$

and the variance of DE is given by

$$\sigma^2 DE \cong \varphi \sum \sigma^2 \hat{DE}_i \quad (i = 1, \ldots, n) \tag{32}$$

where $E(\hat{DE}_i)$ and $\sigma^2 \hat{DE}_i$ are computed for each CSCI from the equations provided in table 1. The Central Limit Theorem [3] implies that for a sufficient number of CSCIs the probability distribution of DE is approximately Gaussian. Therefore, the

22

cumulative probability distribution of DE may be determined from the statistical tables for the normal distribution through the classical transformation

$$z = (DE-E(DE))/\sigma DE$$

where z has a standard normal distribution. An example of how the ASEP model is applied to a system of interrelated CSCIs is given in Case B of Appendix A.

LIST OF REFERENCES

1. Boehm, B.W., "Software Engineering Economics," Englewood
    Cliffs, NJ: Prentice-Hall, 1981.

2. Jensen, R.W., "Comparison of the Jensen and COCOMO Schedule and
    Cost Estimation Models," ISPA Conference Proceedings Vol. III,
    Number 1, San Francisco, CA, 1984.

3. Larson, H.J., "Introduction to Probability Theory and
    Statistical Inference," John Wiley & Sons, Inc., New York,
    1969.

4. Breipohl, A.M., "Probabilistic Systems Analysis," John Wiley &
    Sons, Inc., New York, 1970.

5. Thibodeau, R., "An Evaluation of Software Cost Estimating
    Models," General Research Corporation, RADC-TR-81-144 (AD A104 226),
    June 1981.

MODEL IMPLEMENTATION AND COMPARISON
TO OTHER SOFTWARE EFFORT RISK MODELS

A.1 GENERAL CASES

The MITRE Cost Technical Center has recently developed, in
parallel to this effort, two additional software effort risk
analysis models. They are the Taylor series approach, and the
RISCOMO tool. The Taylor series approach, developed by F.D. Powell[*]
makes use of Taylor series in approximating the mean and the
variance of the development effort. The RISCOMO tool is a
non-analytic model which employs a 1000 sample Monte Carlo
simulation to measure the impact of size uncertainty on effort
estimates. The three cases presented in this appendix compare the
performance of the Taylor series approach, the RISCOMO tool, and the
Analytic Software Effort Probability (ASEP) model. The Constructive
Cost Model (COCOMO) embedded mode (a = 2.8, β = 1.2 in equation 1)
was used to compute the development efforts in each case.

A.1.1 CASE A

This case examines a single hypothetical CSCI with the
technical characteristics given in table A1.

Table A1

Case A Technical Characteristics

CSCI	Size Distribution KDSI	Mean Size KDSI	Πf_j	α
1	Trng (30,50,80)	53.33	1	2.8

[*]The MITRE Corporation

Table A2 provides the development effort distribution, mean, and standard deviation which resulted from implementing the ASEP model, the RISCOMO tool, and the Taylor series approach on the technical parameters of the CSCI given in table A1.

Since this case involved a single CSCI, the ASEP model provided the exact development effort distribution, mean, and standard deviation. The ASEP model results were obtained from equations 15, 16, and 17 (with $\alpha = 2.8$, $\beta = 1.2$) defined in section 2.3. The RISCOMO tool (based on 1000 samples) provided an approximation to the development effort distribution, mean, and standard deviation based on the use of a Monte Carlo simulation method. The Taylor series approach relies on the Central Limit Theorem to contend that the effort distribution is Gaussian. However, since the Central Limit Theorem applies to systems involving more than one CSCI, the Taylor series approach was not used to formulate the development effort distribution for this case, but only to approximate the development effort mean and standard deviation.

Note in table A2 that the difference between the ASEP model and the RISCOMO tool at the 10nth percentile (n = 1, ..., 9) did not exceed 1 percent. At the 95th percentile, there was a 2 percent difference between ASEP and RISCOMO. At the 99th percentile, the difference between ASEP and RISCOMO ranged from 0.4 percent to 5.6 percent.

The RISCOMO tool and the Taylor series method yielded approximations to the development effort mean and standard deviation which evidenced no appreciable difference with the exact development effort mean and standard deviation generated by the ASEP model.

26

Table A2
Case A Development Effort
Cumulative Probability Distribution

Development	Development Effort		
Effort Distribution (Percentile)	ASEP Model	RISCOMO Tool	Taylor Series[1] Approach
10	234.2	231.9	--
20	263.6	268.5	--
30	286.5	288.8	--
40	306.1	309.2	--
50	325.4	325.5	--
60	347.0	349.9	--
70	371.7	370.2	--
80	401.5	402.8	--
90	440.8	439.4	--
95	468.9	459.7	--
99	506.9	480.1-504.5	--
Mean	332.3	334.9	330.8
Standard Deviation	76.6	76.5	76.5

Note: 1. Due to its reliance on the Central Limit Theorem, the Taylor series approach was not used to formulate the development effort distribution in this case.

A.1.2 CASE B

This case examines a 2-CSCI interrelated system with the
technical characteristics given in table A3.

Table A3

Case B Technical Characteristics

CSCI	Size Distribution KDSI	Mean Size KDSI	Development Effort Multipliers			
			RELY	CPLX	DATA	TIME
1	Trng (2,5,10)	5.666	VH	H	*	*
2	Trng (7,9,20)	12.0	VH	*	H	H

Note: 1. These are COCOMO parameter ratings - VH refers to Very
High, H refers to High, and * refers to a Nominal rating.

To use the ASEP model, it was necessary to compute $\hat{DE}_1$ and $\hat{DE}_2$
based on the mean size of each CSCI given in table A3. Applying
COCOMO to the technical parameters of CSCI 1 using the mean size of
5.666 KDSI yielded a development effort $\hat{DE}_1$ of 36.0 staff months.
From equation 1 (with a = 2.8, β = 1.2, and $\bar{I}_1$ = 5.666)

$$\Pi f_{j1} = 1.604$$
$$\alpha_1 = 2.8 \ (\ \Pi f_{j1}) = 4.491$$

where Πf_{j1} is the product of the development effort multipliers for
CSCI 1.

Similarly, applying the COCOMO model to the technical
parameters of CSCI 2 using the mean size of 12 KDSI yielded a
development effort $\hat{DE}_2$ of 94.2 staff months.

28

From equation 1 (with a $\approx$ 2.8, r = 1.2, and $\bar{T}_2$ = 12.0), it may be determined that

$$\Pi f_{j2} = 1.705$$

$$\mu_2 = 2.8 \ (\ \Pi f_{j2}) = 4.775$$

Let DE_T be the COCOMO-generated total development effort of the system based on the mean size of each CSCI, and modeled as an interrelated hierarchy. This was accomplished through the use of an automated version of COCOMO which resides on the MITRE Ultrix system. The resultant DE_T, given the technical parameters in table A3, was 147.6 staff months. From equation 29,

$$147.6 = \lambda [36.0 + 94.2]$$

thus λ = 1.133. Using equations 31 and 32, respectively,

$$E(DE) \cong 1.133[E(\hat{DE}_1) + E(\hat{DE}_2)] = 1.133 \ [36.373 + 94.829] = 148.7$$

$$\sigma^2_{DE} = (1.133)^2 [\sigma^2_{\hat{DE}_1} + \sigma^2_{\hat{DE}_2}] = (1.133)^2[158.9 + 738.124] = 1152.8$$

where $E(\hat{DE}_1)$, $E(\hat{DE}_2)$, $\sigma^2_{\hat{DE}_1}$, $\sigma^2_{\hat{DE}_2}$ were computed from the equations provided in table 1. Therefore, given the Case B technical baseline provided in table A3, the ASEP model approximation of the development effort mean and standard deviation is

$$E(DE) \cong 148.7 \text{ staff months}$$
$$\sigma_{DE} \cong (1152.8)^{1/2} = 33.9 \text{ staff months}$$

Table A4 compares these measures of central tendency with the results generated by the RISCOMO tool and the Taylor series approach.

Table A4

Case B Development Effort

Mean and Variance Estimates

	Development Effort	
	Mean	Standard Deviation
ASEP Model	148.7	33.9
RISCOMO Tool	147.4	33.9
Taylor Series	148.0	33.1

Each model evidenced no appreciable difference in their approximations to the development effort mean and standard deviation.

The RISCOMO tool should be used in this case to formulate the development effort distribution, because it is empirically derived through a Monte Carlo simulation. Since Case B involves a system comprised of just two CSCIs, the Central Limit Theorem (used to contend that the development effort is Gaussian) is not particularly strong for such a small sample. Thus, the ASEP model and the Taylor series approach would not be appropriate in formulating the effort distribution in this case.

Case C describes a 15-CSCI system where the effort distributions formulated by the ASEP model, the RISCOMO tool, and the Taylor series approach may be compared.

A.1.3 CASE C

This case examines a hypothetical system of multiple CSCIs with the technical characteristics given in table A5.

Table A5

Case C Technical Characteristics

CSCI	Size Distribution DSI	Mean Size DSI	μf_j	α_j
1	RtTrng (6437, 8883)	7253	2.268	6.350
2	RtTrng (1287, 1660)	1412	2.449	6.859
3	RtTrng (7750, 9222)	8241	2.474	6.929
4	RtTrng (8525, 12276)	9775	2.124	5.949
5	RtTrng (12231, 18347)	14270	2.278	6.378
6	RtTrng (8000, 14800)	10266	2.401	6.723
7	RtTrng (6975, 9347)	7765	1.406	3.937
8	RtTrng (32650, 40812)	35371	2.420	6.777
9	RtTrng (9000, 11250)	9750	1.904	5.333
10	RtTrng (3219, 4023)	3487	1.524	4.268
11	RtTrng (1288, 1609)	1395	1.515	4.243
12	RtTrng (21700, 31899)	25099	2.586	7.243
13	RtTrng (10075, 10881)	10343	2.401	6.724
14	RtTrng (64375, 67593)	65448	1.265	3.543
15	RtTrng (32188, 33797)	32724	1.265	3.543

Table A6 provides approximations to the development effort distribution, mean, and standard deviation generated by the ASEP model, the RISCOMO tool, and the Taylor series approach.

Table A6

Case C Development Effort Cumulative

Probability Distribution

Development Effort Distribution (Percentile)	Development Effort		
	ASEP Model	RISCOMO Tool	Taylor Series Approach
10	3742	3741	3743
20	3785	3780	3785
30	3816	3806	3815
40	3842	3833	3841
50	3867	3858	3865
60	3892	3885	3889
70	3918	3911	3915
80	3949	3938	3945
90	3992	3990	3987
95	4028	4028	4022
99	4095	4095	4087
Mean	3867	3866	3865
Standard Deviation	98	95	95

Note that the difference between the ASEP model, the RISCOMO tool, and the Taylor series approach at the 10nth percentile ($n = 1$, ..., 9.9) does not exceed 0.3 percent.

A.2 SUMMARY REMARKS

Extending the analytical model to an n-CSCI system provided a method for approximating $E(DE)$, $\sigma^2 DE$, and the development effort distribution while making use of the closed form equations developed in Section 2.

Case A, where the system consisted of a single CSCI, the ASEP model provided the exact values of $E(DE)$, $\sigma^2 DE$, and the development effort distribution. The RISCOMO tool provided an approximation to $E(DE)$, $\sigma^2 DE$, and the development effort distribution through the use of its Monte Carlo simulation algorithm. In Case A, the Taylor series approach also provided good approximations to $E(DE)$, and $\sigma^2 DE$, however, due to its reliance on the Central Limit Theorem in assuming that the development effort distribution is Gaussian, a larger number of CSCIs (more than the single CSCI given in Case A) is necessary in order to use the Taylor series approach to formulate the development effort distribution in this case.

Case B, where the system consisted of just two CSCIs, the ASEP model, the RISCOMO tool, and the Taylor series approach yielded no appreciable difference in their approximations to $E(DE)$ and $\sigma^2 DE$. Since the ASEP model applied to a system of two or more interrelated CSCIs assumes (like the Taylor series approach) that $E(DE)$ and $\sigma^2 DE$ are Gaussian, it should not be used to obtain the development effort distribution for Case B. This case is an example of where the RISCOMO tool is especially useful. Since RISCOMO is based on a Monte Carlo simulation approach, its development effort distribution is empirically derived, hence the size of the sample (i.e., the number of interrelated CSCIs comprising the system) does not affect the RISCOMO formulation of the development effort distribution.

33

Furthermore, in addition to systems consisting of just a few CSCIs, the RISCOMO tool is useful for approximating the development effort distribution for CSCIs with <u>severely</u> skewed size distributions.

Case C consisted of enough CSCIs to show that the performance of the ASEP model, the RISCOMO tool, and the Taylor series approach could be compared. For the scenario given in Case C, the development effort distributions formulated by each model revealed no appreciable difference.

To summarize, the ASEP model has the feature of producing exact values of $E(DE)$, $\sigma^2 DE$, and the development effort distribution given the case of a single CSCI. For systems consisting of many CSCIs, the ASEP model produces approximations to these measures. Its analytic nature makes the ASEP model computationally easy and inexpensive to execute on a computer.

The Taylor series approach is similar to the ASEP model in the sense that it is an analytic model. However, unlike the ASEP model, the Taylor series approach cannot be used to formulate the development effort distribution of a single CSCI (as in Case A) due to its reliance on the Central Limit Theorem. By reason of its analytic nature, the Taylor series approach also has the feature of being computationally inexpensive in computer usage.

The RISCOMO tool can be applied to a broad set of scenarios, however, it is not computationally straightforward, and can become expensive in computer usage due to its Monte Carlo simulation nature. RISCOMO is especially useful in cases involving a few CSCIs and in circumstances where the size distributions of a system of CSCIs are <u>severely</u> skewed.

DATE
FILMED
3 8